



US009880995B2

(12) **United States Patent**  
**Chaski**

(10) **Patent No.:** US 9,880,995 B2  
(45) **Date of Patent:** Jan. 30, 2018

(54) **VARIABLES AND METHOD FOR  
AUTHORSHIP ATTRIBUTION**

(76) Inventor: **Carole E. Chaski**, Georgetown, DE  
(US)

(\*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 984 days.

(21) Appl. No.: **11/398,728**

(22) Filed: **Apr. 6, 2006**

(65) **Prior Publication Data**

US 2007/0239433 A1 Oct. 11, 2007

(51) **Int. Cl.**

**G06F 17/27** (2006.01)

**G06F 17/30** (2006.01)

(52) **U.S. Cl.**

CPC ..... **G06F 17/27** (2013.01); **G06F 17/271** (2013.01); **G06F 17/2705** (2013.01); **G06F 17/2765** (2013.01); **G06F 17/2785** (2013.01); **G06F 17/30707** (2013.01)

(58) **Field of Classification Search**

CPC .... G06F 17/27; G06F 17/2705; G06F 17/271;  
G06F 17/2765; G06F 17/2785

USPC ..... 704/1–10

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,182,708 A \* 1/1993 Ejiri ..... 704/1  
5,799,268 A \* 8/1998 Boguraev ..... 704/9  
6,473,752 B1 \* 10/2002 Fleming, III ..... 707/4  
6,473,753 B1 \* 10/2002 Katariya et al. ..... 707/4  
6,640,145 B2 \* 10/2003 Hoffberg et al. ..... 700/83  
6,963,839 B1 \* 11/2005 Ostermann et al. ..... 704/260

7,088,949 B2 \* 8/2006 Burstein et al. ..... 434/353  
7,137,070 B2 \* 11/2006 Brown et al. ..... 715/744  
7,233,891 B2 \* 6/2007 Bond et al. ..... 704/9  
7,296,015 B2 \* 11/2007 Poltorak ..... 707/5  
2002/0052730 A1 \* 5/2002 Nakao ..... 704/10  
2004/0078192 A1 \* 4/2004 Poltorak ..... 704/9  
2007/0260448 A1 \* 11/2007 Lorenzen et al. ..... 704/1

OTHER PUBLICATIONS

Kim Luyckx, Walter Daelemans, "Shallow Text Analysis and Machine Learning for Authorship Attribution", Proceedings of the Fifteenth Meeting of Computational Linguistics in the Netherlands, 2005.\*

E. Stamatatos, N. Fakotakis, G. Kokkinakis, "Computer-Based Authorship Attribution Without Lexical Measures", Computers and the Humanities, 2001.\*

H. Baayen, H. van Halteren, A. Neijt, F. Tweedie, "An Experiment in Authorship Attribution", Journes internationales d'analyse statistique des Donnes Textuelles 6, 2002.\*

(Continued)

*Primary Examiner* — Eric Yen

(74) *Attorney, Agent, or Firm* — James M. Lennon

(57) **ABSTRACT**

A method uses linguistic units of analysis to identify the authorship of a document. The method is useful to determine authorship of brief documents, and in situations where there are less than ten documents per known author, i.e. when there is scarcity of text. The method analyzes parameters such as the syntax, punctuation, and, optionally the average word and paragraph length, and when the parameters are analyzed using statistical methods, obtains a high degree of reliability (>90% accuracy). The method can be applicable to numerous languages other than English because the variables selected are characteristic of most languages. The reliability of the method is verified when subjected to a cross-validation statistical analysis.

**2 Claims, 5 Drawing Sheets**

